Using Orthogonal Locality Preserving Projections to Find Dominant Features for Classifying Retinal Blood Vessels

Relan, Devanjali; Ballerini, Lucia; Trucco, Emanuele; MacGillivray, Tom

Published in:
Multimedia Tools and Applications

DOI:
10.1007/s11042-018-6474-7

Publication date:
2019

Document Version
Peer reviewed version

Noname manuscript No.
(will be inserted by the editor)

Using Orthogonal Locality Preserving Projections to Find Dominant Features for Classifying Retinal Blood Vessels

Devanjali Relan · Lucia Ballerini · Emanuele Trucco · Tom MacGillivray

Received: date / Accepted: date

Abstract Automatically classifying retinal blood vessels appearing in fundus camera imaging into arterioles and venules can be problematic due to variations between people as well as in image quality, contrast and brightness. Using the most dominant features for retinal vessel types in each image rather than predefining the set of characteristic features prior to classification may achieve better performance. In this paper, we present a novel approach to classifying retinal vessels extracted from fundus camera images which combines an Orthogonal Locality Preserving Projections for feature extraction and a Gaussian Mixture Model with Expectation-Maximization unsupervised classifier. The classification rate with 47 features (the largest dimension tested) using OLPP on our own ORCADES dataset and the publicly available DRIVE dataset was 90.56% and 86.7% respectively.

Keywords Retina · Fundus images · Vessel classification · Feature extraction · Orthogonal Locality Preserving Projections (OLPP)

Devanjali Relan
Centre for Clinical Brain Sciences, University of Edinburgh, U.K.
E-mail: devanjali_relan@ed-alumni.net

Lucia Ballerini
Centre for Clinical Brain Sciences, University of Edinburgh, U.K.
E-mail: lucia.ballerini@ed.ac.uk

Emanuele Trucco
Computing, School of Science and Engineering, University of Dundee, U.K.
E-mail: E.Trucco@dundee.ac.uk

Tom MacGillivray
Centre for Clinical Brain Sciences, University of Edinburgh, U.K.
E-mail: T.J.MacGillivray@ed.ac.uk

1 Introduction

Quantitative structural analysis of the retinal blood vessels as they appear in fundus camera imaging provides important indicators for diseases which affect the body and brain such as diabetes, hypertension and Alzheimers disease and which manifest in the retina [1–3]. The quantifiable vascular changes that appear can be different for arterioles and venules. For example, narrowed retinal arterioles are associated with long-term risk of hypertension [2], while larger retinal venular calibre has been associated with Alzheimer's disease [3]. Thus, an essential part of a computerised system for retinal vascular characterisation is an automatic technique for discerning arterioles from venules.

To date, several techniques have been developed and introduced to identify vessel type using various discriminative features with different supervised [4–13] and unsupervised algorithm [14–16]. Colour, location, patterning and structure are commonly used to discriminate between vessel type [8]. Classification performance not only varied with the choice of features used and the classification algorithm employed but also on the overall classification framework.

The unsupervised classification approach of [14,15] has a drawback, that it imposes a condition to have at least one vein and one artery per quadrant. Moreover, basic k-means clustering approach is sensitive to initialisation and therefore, can easily get stuck at local minima. The authors in [4–13,17] used a supervised approach which requires large volumes of clinical annotations (i.e. manual labelling of vessels into venules and arterioles) to generate the requisite training data and this may not be easy to source, especially for large dataset.

The authors in [4,5,10] used different feature selection methods such as sequential forward floating selection (SFFS), a wrapper based method to find the discriminative features for differentiating between arterioles and venules. SFFS starts from the empty set, and the next best single feature is added. After each forward step, SFFS performs backward steps to exclude any feature that becomes non-useful. Using this method near optimality is achieved, but at the expense of computational time especially in the case of data of greater complexity and dimensionality [18]. For large dataset, it may not be a very good approach.

The authors in [13] provided the extensive study using WEKA (Waikato Environment for Knowledge Analysis) software [19] for retinal vessel classification using nine different feature extraction and thirteen machine learning algorithms (both single and ensemble classifiers). They found that the information extracted by Local Binary Patterns (LBP) and Multiscale Rotation Invariant LBP (MS-RI LBP) from the arteries and veins vessels are most discriminative features. The performance of the proposed method using the LBP and MS-RI LBP features was tested with 13 different classifiers. Out of the 13 tested classifiers, only 6 methods gave the result greater than 85%. This means that to obtain high classification rate, not only the features but also the classifier and overall framework is equally important.

Furthermore, in [17], the authors proposed a semi-automatic system for retinal vasculature classification into the artery and vein. They extracted colour and structural features from segmented vessels of DRIVE dataset and then utilised PCA to reduce the dimension of the feature space. Finally, the neural network classifier was used to classify the vessels. They divided the total 69 extracted features into 4 set based on the eigenvalues. They showed that different feature set yielded different accuracy and highest accuracy was obtained with features corresponding to highest eigenvalues.

Similarly, Hausdorff distance measure has been used extensively in computer vision applications for feature selection as well as to determine the similarity of patterns [20–23]. Moreover, recently the results obtained by authors in [24] with correlative classifiers approach based on particle filter seems to be very promising. As mentioned above, representative features as well as robust classifier play an important role for an accurate retinal classification system. For a comprehensive study of several proposed methods for classification of arteries-veins classification in fundus images the readers are referred to [25]. Moreover, a recent exhaustive review study related to latest progress in image processing techniques in feature-based retinal image analysis can be found here [26]

2 Contribution and Organisation of the Paper

Automated vessel classification relies on representative features extracted from the images. Predefining the feature set across all images is problematic due to variations in colour as well as the intensity within and between the images [5, 6]. Additionally, the image resolution as well as the image quality may not be constant throughout an image set due to different applications, operators and camera systems. Hence, extracting the most dominant features in each image may lead to a better classification performance.

Therefore, in this paper, we propose a classification framework for arteriole-venule (a-v) classification based on a discriminative set of features that are extracted using Orthogonal Locality Preserving Projections (OLPP) [27], and vessel classification is subsequently performed using Gaussian Mixture Model with Expectation-Maximization (GMM-EM) unsupervised classifier (without the need of training data). We also studied the effect of feature set dimension on the performance of the proposed hybrid approach (OLPP + GMM-EM).

The paper is structured as follows: Section 3 provides the datasets that were used to test the proposed vessel classification approach. Section 4 describes the methodology and the main steps involved in the retinal vessel classification. Section 5 presents the results obtained, which are then discussed in Section 6 and followed by conclusions in Section 7.

3 Materials

The proposed vessel classification method was evaluated on our own dataset: ORCADES (Orkney Complex Disease Study) and a publicly available database, the DRIVE (Digital Retinal Image for Vessel Extraction),

3.1 Datasets

1. ORCADES: Seventy colour fundus camera images were selected at random from the Orkney Complex Disease Study (ORCADES) large image database containing $\approx$ 2000 images. The ORCADES study is a family-based, cross sectional and genetic epidemiology study based on an isolated population in the north of Scotland that aims to discover the genes and their variants which influence the risk of common, complex diseases[28]. Each image in the database had the resolution of 2048×3072 pixels and was captured with Canon CR-DGi non-mydriatic retinal camera with 45° field of view (FoV). From the 70 images, 802 vessels from zone B (an annulus 0.5 to 1 OD diameter from the OD boundary) were extracted to test the vessel classification framework.
2. DRIVE: The DRIVE dataset contains 40 colour fundus camera images [29], each having the resolution of 768×584 pixels. The set of 40 images has been divided into a training and a test set, both containing 20 images each. The retinal photographs were obtained from a diabetic retinopathy screening program in the Netherlands and were acquired using a Cannon CR5 non-mydriatic 3 CCD camera with a 45° field of view (FoV). In this study, 171 vessels extracted from zone B of 20 test images are used to test the performance of the proposed method.

3.2 Manual labelling

The manual labelling of the vessels was undertaken by two trained human observers (authors; DR and TM). Both observers were involved in retinal imaging and analysis for clinical research. Observer 1 (DR) classified all vessels into arterioles and venules, while 1.5% of 802 vessels of ORCADES images (i.e. 12 vessels) were not classified (labelled as unclassified) by observer 2 (TM) due to an uncertainty in deciding the label. Out of all the vessels which were classified (as arterioles or venules) by both observers (i.e. 790), it was observed that the discrepancy in labelling vessels by two observers were found for only two vessels i.e. the labels given by both observers were same for 788 vessels.

3.3 Performance measures

Following performance measures [30] were computed for observer 1 (separately for arterioles and venules) w.r.t. the labels of observer 2,

$$\text{Sensitivity} = \frac{TP_{a/v}}{(TP_{a/v} + FN_{a/v})} \tag{1}$$

$$\text{Specificity} = \frac{TN_{a/v}}{(TN_{a/v} + FP_{a/v})} \tag{2}$$

$$\text{Positive Predicted Value} = \frac{TP_{a/v}}{(TP_{a/v} + FP_{a/v})} \tag{3}$$

$$\text{Negative Predicted Value} = \frac{TN_{a/v}}{(TN_{a/v} + FN_{a/v})} \tag{4}$$

$$\text{Positive Likelihood Ratio} = \frac{Sensitivity}{(1 - specificity)} \tag{5}$$

$$\text{Negative Likelihood Ratio} = \frac{(1 - Sensitivity)}{specificity} \tag{6}$$

$$\text{Classification Accuracy} = \frac{(TP_{a/v} + TN_{a/v})}{(TP_{a/v} + TN_{a/v} + FP_{a/v} + FN_{a/v})} \tag{7}$$

$$\text{Classification Error Rate} = \frac{(FP_{a/v} + FN_{a/v})}{(TP_{a/v} + TN_{a/v} + FP_{a/v} + FN_{a/v})} \tag{8}$$

where *TP* is *True Positive*, *FP* is *False Positive*, *TN* is *True Negative* and *FN* is *False Negative*. These are further defined as [15]:

- True Positive$_{\mathbf{a}rtery(\mathbf{v}ein)}$ ($TP_{a/v}$): When both the observer 1 and the observer 2 identified a vessel as arteriole (or venule).
- False Positive$_{\mathbf{a}rtery(\mathbf{v}ein)}$ ($FP_{a/v}$): When the observer 1 identified a vessel as arteriole (or venule) and the observer 2 identified it as venule (or arteriole).
- True Negative$_{\mathbf{a}rtery(\mathbf{v}ein)}$ ($TN_{a/v}$): When the observer 2 identified a vessel as venule(arteriole) and the observer 1 identified it as not arteriole (venule)(i.e. either as venule (arteriole) or not labeled).
- False Negative$_{\mathbf{a}rtery(\mathbf{v}ein)}$ ($FN_{a/v)}$): When the observer 1 did not identify a vessel as arteriole (or venule), i.e. either as venule(arteriole) or not labeled, and the observer 2 identified it as arteriole (venule).

The performance measures are given in Table 1. This demonstrates the high level of agreement between the two observers. Thus, the results in the subsequent section are given with respect to (w.r.t.) observer 1.

Table 1: Table showing the performance of observer 1 w.r.t. observer 2 who both manually labelled vessels as arterioles or venules on the ORCADES dataset (70 images; 802 vessels).

Performance Measure	**arterioles**	**venules**
Sensitivity	0.9976	0.9973
Specificity	0.9973	0.9976
Positive Predicted value	0.9976	0.9973
Negative Predicted value	0.9973	0.9976
Positive Likelihood Ratio	367.13	420.85
Negative Likelihood Ratio	0.0024	0.0027
Classification Accuracy	0.9975	0.9975
Classification Error rate	0.0025	0.0025

4 Methodology

4.1 Pre-processing

First a fundus camera image was pre-processed to compensate for variation in brightness and contrast caused by non-uniform background illumination. This was enacted upon multiple channels utilized for obtaining features for vessel classification - **R**ed, **G**reen, and **B**lue from **RGB** image space; **H**ue and **V**alue from **HSV** image space; Lightness from the **L**ab image space; and a gray scale image, Gy, converted from RGB (where $Gy = 0.299\times\mathbf{R}+0.587\times\mathbf{G}+0.114\times\mathbf{B}$). The pre-processing used a correction technique based on median filtering [31].

To do this, first, background intensity, $BI = BI(i,j)$, in each channel was estimated by median filtering with a mask of size 100×100. The mask size was chosen in such a way that it is several times bigger than the expected maximum diameter of the retinal vessels. Then the correction coefficients, $CC(i,j)$, were calculated by dividing the maximum gray-level value, $max(BI)$, in the median filtered image by the intensity level value of each pixel in the filtered image,

$$CC(i,j) = \begin{cases} \frac{max(BI)}{BI(i,j)} & \text{for } BI(i,j) \neq 0 \\ 0 & \text{for } BI(i,j) = 0 \end{cases} \tag{9}$$

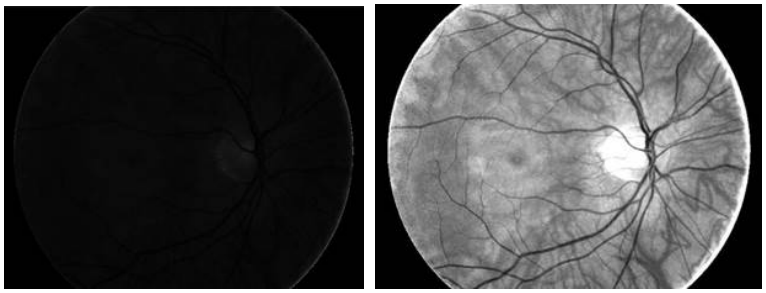
where (i,j) represent image coordinates. The corrected output image, $OI(i,j)$, is then obtained by multiplying the input image, IM, with its correction coefficients,

$$OI(i,j) = IM(i,j) \times CC(i,j) - c \tag{10}$$

where c $= max(BI) - 128$ is an image dependent constant and is subtracted from whole image. By subtracting the constant c, the mean value is moved to the middle of the gray-value range between 0 to 255.

In addition, the hue channel was pre-processed, prior to background correction, to improve the contrast of vessels against background by mapping the

original pixel intensity values between 1% of bottom and 20% of top pixel intensity values to values between 0 and 1 respectively. Fig. 1 shows an example of the Hue channel of a fundus image before and after the contrast adjustment.



(a) (b)

Fig. 1: H channel (a) before and, (b) after contrast adjustment.

4.2 Detecting centreline pixels

To extract vessel centreline pixels, the green channel of a fundus camera image was divided into four quadrants by locating the OD and its approximate diameter with software described in [32] (see Fig. 2(a), solid black lines passing through OD centre). In zone B (i.e. the region between the two concentric circles in Fig. 2(a)) each vessel was tracked between two manually marked start and end points, see Fig. 2(b) [30]. The cross-sectional intensity profiles were found at every 5^{th} pixel.

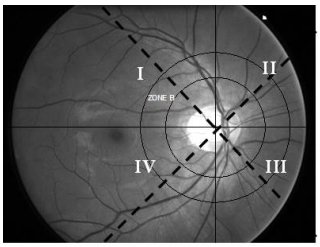
To do this, first, the start (S) and end (E), were marked manually on the vessels between which the vessel needs to be tracked (see Fig. 2(b)). Then the vector defining the direction from S to E, V_{S-E}, was computed as $V_{S-E} = [V_{xx}\ V_{yy}]$, where $V_{xx} = d_e\cos(\theta)$ and $V_{yy} = d_e\sin(\theta)$, with d_e as the Euclidean distance between S and E and $\theta = tan^{-1}(d_{yy}/d_{xx})$, where d_{xx} and d_{yy} are differences between x and y coordinates of points E and S. Then coordinates of the new point, P_{new}, 5 pixels ahead of S was calculated as,

$$P_{new} = S + 5(V_{S-E}). \tag{11}$$

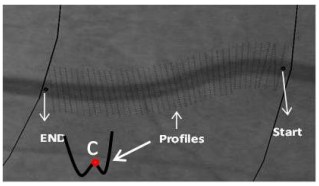
At P_{new}, the intensity profile across the vessel axis (and which resembles an inverted Gaussian; see Fig. 2(b) which shows an example) was obtained. Point C, marked red on the intensity profile (see Fig. 2(b)), give the approximate centre of the vessel and was found by locating and averaging 2 local minima on the profile. Then, the next P_{new} was calculated (using eq. 11) five pixels ahead of point C with vector direction V_{C-E}. This procedure continues until end point E was reached, i.e. cross-sectional intensity profiles (black straight lines in Fig. 2(b)) were found at every 5^{th} pixel between S and E. Then

the Canny edge detector [33] was applied to each of these intensity profiles, to locate vessel edges (marked as black circles in Fig. 2(b)) and finally the centreline pixels (marked as green in Fig.2(c)) were located as the midpoint of a pair of edge points.

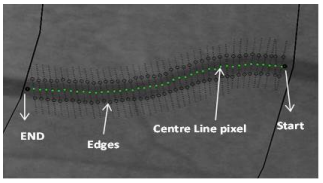
In this way, the centreline pixels were extracted from vessels in each quadrant (i.e. from quadrant I, II, III and IV; see Fig. 2(a)). The centreline pixels were extracted from vessels in each quadrant to provide a set of v vessel segments $S_q = V_{1q}, V_{2q},...,V_{vq}$, where V_{iq} is i^{th} vessel in q^{th} quadrant, S_q is set of vessels in q^{th} quadrant and each vessel is represented by a set of h centreline pixels: $V_{iq} = p_{1q}, p_{2q},...,p_{hq}$. Moreover, the distance between the edge points were stored as an approximate vessel width to evaluate features inside the vessels.



(a)



(b)



(c)

Fig. 2: (a) Figure shows normal four quadrants i.e. quadrant I, II, III and IV, and when they were rotated by 45 ° (marked as dashed line in the image), (b) Profiles between start and end points, (c) Centreline pixel extraction on each of the profile using canny edge detector.

4.3 Extracting features

Once the centreline pixels were found in each quadrant, features were extracted by sampling inside the vessels in each of the pre-processed channels. The details of this are as follows:

1. **ROI-based** features were extracted using a circular neighborhood around each centreline pixel, with diameter 60% of the mean vessel diameter, to calculate the intensity level inside the vessel.

2. **Profile-based** features were extracted from the background corrected **R** and **G** channels as these channels show good contrast discrimination between arterioles and venules [34,35,14],) from the **RG**B colour space after an additional contrast adjustment (by mapping the intensity values in an input image to new values such that 1% of data is saturated at low and high intensities value of input image) from a profile line across the vessel of length 60% of the mean vessel diameter.
3. **Contrast** of a vessel w.r.t. background from the corrected **R** and **G** channels was extracted after additional contrast adjustment in a similar manner as explained above for profile based feature. For each centreline pixel, a profile of length 2.5 times the width of the vessel was drawn. The profile length was chosen bigger than the twice the approximate width of the vessel to calculate the background intensity at each end of the profile. At each end of the profile, mean intensity I_{b1} and I_{b2}, from two circular ROI's (whose diameters were 60% of the mean vessel diameter, and a centre as the end point of profile) were obtained as shown in Fig. 3.
   The average of these two intensities, I_b, was calculated and this represented the background intensity. The mean intensity I from a circular neighbourhood around the centreline pixel (with diameter 60% of the mean vessel diameter) was extracted as the vessel intensity. The contrast feature, C, was then calculated as,

$$C = \frac{I - I_b}{I_b} \tag{12}$$

For ROI and profile-based features, mean *(M)*, standard deviation *(Std)*, variance *(Var)*, minimum *(Min)* and maximum *(Max)* of intensity values from each of the channels were extracted. This along with contrast features created a set of 47 features for each image. See Table 2 which summarizes the feature set.

In order to check the consistency, robustness and influence of the dimension or size of the feature set on vessel classification, feature subsets of different dimensions were formed from the set of 47 features. See Table 3 for a summary. In the context of this paper, a feature dimension of ≤ 15 was considered small whereas > 15 was considered as high. The feature subsets with smaller dimensions were formed using ROI based colour features, which are commonly reported as discriminant features to classify retinal vessel type [9,5,7] - i.e. those extracted from R, G, B, H and V. Whereas, the feature sets with larger dimensions were formed with ROI-based colour, profile-based and contrast based features from R, G, B, H, V, L and Gy.

In RGB colour fundus images, the green and red channels usually show good contrast discrimination between arterioles and venules, and therefore the feature set with dimension 9 was formed by combining the features extracted from the red and green channels with those extracted from **B**, **H** and **V** respectively. These feature sets are referred to as 9**B**, 9**H** and 9**V** henceforth, respectively (see Table 3). OLPP was then used to extract the most dominant features for an image from these feature subsets.

Table 2: Description of Total Number (No.) of features extracted from each image

Features	No. of features	Channels	No. of channels	Total No. features
ROI-based: M, Std, Var, Min, Max	5	G , R, B, H, V, L, Gy	7	35
Profile-based: M, Std, Var, Min, Max	5	G, R	2	10
Contrast	1	G, R	2	2

Table 3: Different Feature sets consisting different colour features extracted from R,G,B,H,V, L channels and grey image.

	Total Features	Features Set
9	**ROI based:** GRB (**9B**)	**ROI based features:** M, Std., Var. from G, R, and B channel
	ROI based: GRH (**9H**)	**ROI based features:** M, Std., Var. from G, R, and H channel
	ROI based: GRV (**9V**)	**ROI based features**: M, Std., Var. from G, R, and V channel
12	**ROI based:** GRBH	**ROI based features:** M, Std., Var. from G, R, B and H channel
15	**ROI based:** GRBHV	**ROI based features:** M, Std., Var. from G, R, B, H and V channel
32	**ROI based:** GRBHVL **Contrast:** G and R channel	**ROI based features:** M, Std, Var, Min, Max from G, R, B, H, ,V and L channel + **Contrast** from G and R channel
37	**ROI based:** GRBHVLGy **Contrast:** G and R channel	**ROI based features:** M, Std, Var, Min, Max from G, R, B, H, V, L channel and Gy image + **Contrast** from G and R channel
42	**ROI based:** GRBHVL **Profile based:** G and R channel **Contrast:** G and R channel	**ROI based features:** M, Std, Var, Min, Max from G, R, B, H ,V and L channel + **Profile based features:** M, Std, Var, Min, Max from G and R + **Contrast** from G and R channel
47	**ROI based:** GRBHVLGy **Profile based:** G and R channel **Contrast:** G and R channel	**ROI based features:** M, Std, Var, Min, Max from G, R, B, H ,V and L channel and Gy image + **Profile based features:** M, Std, Var, Min, Max from G and R + **Contrast** from G and R channel

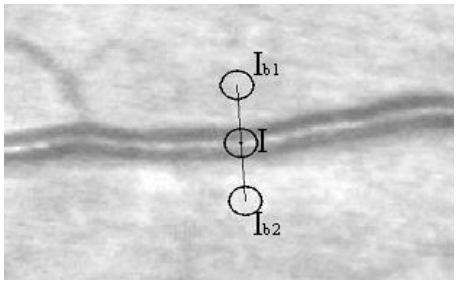


Fig. 3: Extracting contrast features w.r.t. retinal background. I is the mean intensity extracted from circular ROI inside the vessels while I_{b1} and I_{b2} are the mean intensity extracted from circular ROI from the vessel background on either side of the vessel.

4.4 Feature extraction using Orthogonal Locality Preserving Projections (OLPP)

When high dimensional data are available, then it is often advisable to find the most informative features before attempting classification. We propose to use Orthogonal Locality Preserving Projections (OLPP) algorithm [27], the to extract the discriminant features to solve the retinal vessel classification problem.

OLPP finds an orthogonal mapping of a given feature set in order to preserve the local structure. OLPP is an extension of Locality Preserving Projections (LPP) [27,36]. LPP preserves the local structure and is nonorthogonal. On the other hand, OLPP produces orthogonal basis functions and therefore has more locality preserving power than LPP. Also, OLPP has a neighbourhood preserving property and can, therefore, better capture the intrinsic manifold structure to a greater extent [27]. The locality preserving ability is directly related to the discriminating ability [27,37]. In contrast, conventional algorithms such as PCA only model the features in Euclidean space and cannot detect the intrinsic low-dimensionality features. The algorithmic procedure of OLPP is as follows [27] [37],

I. **PCA Projection**: Let $f_1, f_2, ..., f_n \subset \mathbb{R}^m$ be a set of features such that $F = [f_1, f_2, ..., f_n]$ forms the feature matrix. In many classification problems, the dimension of the features m is often much larger than the number of features n. In order to overcome this problem, PCA projects the features set into a subspace by removing the eigenvectors corresponding to zero eigenvalue. Thus the extracted features are statistically uncorrelated. The transformation matrix of PCA is denoted by W_{PCA}.
II. **Adjacency Graph construction**: Let U be a J nearest neighbour graph with n nodes. The i-th node corresponds to the feature f_i. If f_i and f_j are close, i.e. f_i is among p nearest neighbours of f_j or f_j is among p nearest neighbours of f_i then an edge is put between nodes i and j. In an unsupervised mode, edges are located between a sample and its J nearest neighbours, where J is a small integer. In order to measure the closeness between two arbitrary data nodes in a J nearest neighbour graph Euclidean distance was used. In an unsupervised mode, the J closest neighbours of one node can be obtained by analysing the constructed the Euclidean distance matrix. The constructed nearest neighbours graph approximates the local manifold structure.
III. **Choosing the Weights**: If the nodes i and j are linked then assign weight as,

$$W_{ij} = e^{-\frac{||f_i - f_j||^2}{2t}} \tag{13}$$

where t is a constant. Otherwise, $W_{ij} = 0$. The weight matrix W of graph G represents the local structure of the feature space.

IV. **Computing the Orthogonal Basis Functions**: Define diagonal matrix D as the sums of each column (or row as W is symmetric) entries of W i.e. $D_{ij} = \sum_j W_{ij}$. Also define Laplacian matrix, $L = D - W$ [27]. Also state,

$$A^{(d-1)} = [\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_{(d-1)}}] \tag{14}$$

$$B^{(d-1)} = [A^{(d-1)}]^T.(FDF^T)^{-1}.A^{(d-1)} \tag{15}$$

where $\{\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_d}\}$ are the orthogonal locality preserving projections and can be computed iteratively as,

(a) Compute $\mathbf{a_1}$ which is the eigenvector of $(FDF^T)^{-1}FLF^T$ corresponding to the smallest eigenvalue.
(b) Compute $\mathbf{a_d}$ as the eigenvector of

$$M^{(d)} = \{I - (FDF^T)^{-1}.A^{(d-1)}[B^{(d-1)}]^{-1}[A^{(d-1)}]^T .(FDF^T)^{-1}.(FLF^T)\} \tag{16}$$

corresponding to the smallest eigenvalue of $M^{(d)}$.

V. **OLPP Embedding**: Let $W_{OLPI} = [a_1, a_2, \ldots, a_l]$, the embedding is defined as,

$$\mathbf{f} \rightarrow \mathbf{y} = W^T\mathbf{f} \tag{17}$$

$$W = W_{PCA}W_{OLPI} \tag{18}$$

where $\mathbf{y}$ is a l-dimensional feature vector $\mathbf{f}$ and $\mathbf{W}$ is the transformation matrix.

4.5 Gaussian Mixture Model with Expectation-Maximisation (GMM-EM)

After feature extraction, the centreline pixels were classified in an unsupervised manner using the Gaussian mixture model expectation-maximisation (GMM-EM) method [38,39].

4.5.1 Gaussian Mixture Model (GMM)

Considering data as a mixture of Gaussian distributions is a widely used means to cluster the data [40]. A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of K Gaussian component densities known as the mode of the GMM, and each mode has its own mean and covariance [38]. Mathematically, For a D-dimensional vector X, the multivariate Gaussian distribution $p(X)$ can be defined as [39],

$$p(X) = \mathcal{N}(X|\mu, \Sigma) \tag{19}$$

where μ is the mean, Σ is the covariance matrix and $\mathcal{N}$ signifies the multivariate Gaussian and can be mathematically defined as follows,

$$\mathcal{N}(X|\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^D|\Sigma|}}\exp\big(-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\big) \tag{20}$$

Similarly the mixture of multivariate Gaussian densities is defined as [39],

$$p(X) = \sum_{k=1}^{K}\pi_k\mathcal{N}(X|\mu_k,\Sigma_k). \tag{21}$$

where, $k = 1\cdots K$, is the number of Gaussian, $0 \leq \pi_k \leq 1$ is the mixture coefficient of k^{th} Gaussian and $\sum_{k=1}^{K}\pi_k = 1$, μ_k is the mean of the k^{th} Gaussian and Σ_k is the covariance matrix of the k^{th} Gaussian respectively. Given a GMM, the main aim is to maximise the likelihood function w.r.t. the parameters (μ_k, Σ_k and π_k) [39]. As compared to conventional clustering method such as K-mean clustering, GMM cluster assignment is much more flexible, i.e. data point belongs to each cluster to a different degree (with different probabilities) as opposed to a hard assignment of a data point to each cluster.

4.5.2 The Expectation-Maximisation (EM) algorithm

Two commonly used methods for inferring the parameters such as means, variances and mixture weights of the unknown distributions from an observed data set are maximum-likelihood (ML) estimation and Bayesian posterior inference. In this paper, we utilise the ML estimation method for estimating the parameters of the unknown distribution. In practice, ML estimation of the unknown densities is generally done using the Expectation-maximisation (EM) algorithm as it is a reliable and faster iterative technique for computing the maximum-likelihood estimate of the parameters of an underlying distribution from data, where a closed form analytical expression is difficult to obtain. The advantage of EM algorithm is that it can also handle efficiently the data which is incomplete or has missing values [39].

The EM algorithm is a local optimisation method, and hence it is sensitive to the initialisation of the model. Therefore, the simplest way to initiate parameters is to use a *K-means* approach [39]. EM for Gaussian mixtures algorithm [39] is explained as,

I. **Initialise Gaussian parameters**: Initialise μ_k , Σ_k and π_k for each Gaussian k using the *K-means* approach and evaluate the initial value of the log likelihood.
II. **E Step**: Calculate an assignment score $\gamma(Z_{nk})$ with the current parameters values,

$$\gamma(Z_{nk}) = \frac{\pi_k\mathcal{N}(X_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j\mathcal{N}(X_n|\mu_j,\Sigma_j)}. \tag{22}$$

Here $n = 1\cdots,N$ is total number of data points.

III. **M Step**: Given scores, adjust μ_k , Σ_k and π_k for each cluster k i.e. for each Gaussian k, update parameters using new $\gamma(Z_{nk})$. Mean of Gaussian k is given by,

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(Z_{nk}) X_n \tag{23}$$

where N_k is defined as the effective number of points assigned to cluster k and given by,

$$N_k = \sum_{n=1}^{N} \gamma(Z_{nk}). \tag{24}$$

Find the means that fits the assignment score best. Covariance matrix of Gaussian k is given by,

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(Z_{nk})(X_n - \mu_k^{new})(X_n - \mu_k^{new})^T. \tag{25}$$

Mixing coefficient for Gaussian k is given by,

$$\pi_k^{new} = \frac{N_k}{N}. \tag{26}$$

IV. **Evaluate log Likelihood**: Calculate the log likelihood again,

$$ln\ p(X|\mu, \Sigma, \pi) = \sum_{n=1}^{N} ln \sum_{k=1}^{K} \pi_k N(X_n|\mu_k, \Sigma_k) \tag{27}$$

and finally check for convergence of either the parameters or the log likelihood. If likelihood or parameters converge then stop else go to step II. (E step).

4.6 Vessel Classification

Each of the centreline pixels belonging to vessels from pair of adjacent quadrants (i.e. from quadrant combinations (I, II), (II, III), (III, IV) and (IV, I)) were classified using GMM-EM with dominant discriminative features provided by OLPP. The centroid of arteries and veins cluster is associated with a vector of mean values representing the features. The two mean intensity values representing the centroids (i.e. for two clusters) were compared to determine the a-v class. The cluster with higher mean intensity value at its centroid is defined as arteriole cluster and the other as venule cluster [16].

As each quadrant was considered twice in the processing for pairs (I, II), (II, III), (III, IV) and (IV, I), each pixel has two labels assigned to it in total [30]. Then the quadrant partitioning was rotated 45 ° clockwise (see Fig. 2(a), dashed black lines at 45 ° to horizontal passing through OD centre) and the centreline pixels belonging to vessels from a pair of adjacent rotated quadrant were classified again, generating two more labels per centreline pixel. In this

way, each centreline pixel has a total of four labels. The vessels were assigned a final status of based on the maximum polling of labels assigned to its centreline pixels.

4.7 Classification Performance Evaluation

The classification rate was calculated in order to compare the performance of the different methods. The classification rate, CR, was calculated as,

$$CR = \frac{\text{CV}}{\text{TV} - \text{TU}} \times 100 \tag{28}$$

where CV is correctly classified vessels, TV is total number of vessels and TU is number of vessels remained unclassified. The performance measure were also calculated in similar manner as explained in section 3 where TP, FP, TN and FN are defined as:

- $TP_{a/v}$: When both the system and the observer identified a vessel as arteriole (or venule).
- $FP_{a/v}$: When the system identified a vessel as arteriole (or venule) and the observer identified it as venule (or arteriole).
- $TN_{a/v}$: When the observer identified a vessel as venule(arteriole) and the system identified it as not arteriole (venule)(i.e. either as venule (arteriole) or not labelled).
- $FN_{a/v)}$: When the system did not identify a vessel as arteriole (or venule), i.e. either as venule(arteriole) or not labelled, and the observer identified it as arteriole (venule).

5 Result

In certain data analysis techniques, high-dimensional data can be replaced by its projection onto the most important axes, and these axes are the ones corresponding to the largest Eigen values. Thus the original data is approximated by data with fewer dimensions (by choosing the most dominant Eigen values), which summarizes the original data well [41]. OLPP fall into this class of data analysis techniques, therefore, eigenvectors corresponding to high eigenvalues were extracted for generating a new feature set. With OLPP, transformed features were used for vessel classification.

In order to select the dominant eigenvalues and corresponding eigenvectors for OLPP, a threshold value was selected by conducting an initial experiment with 10 images. Based on the analysis, in order to select the most important features (corresponding to higher Eigen values), a cut-off of 0.9 and 0.95 were used for smaller and larger dimension of feature set, respectively. Thus with OLPP, the Eigen vectors corresponding to Eigen values greater than or equal to 0.9 (for smaller dimensions) and 0.95 (at larger dimensions) were used to transform the original feature set.

Table 4 presents the rate of correct classification rate (CR), and unclassified vessels (UN) obtained using the proposed method from ORCADES dataset. The mean and standard deviation of the classification results obtained with smaller (9-15 features) as well as larger dimensions (32-47 features) were calculated and are also mentioned in Table 4.

Table 4: Correct Classification Rate (CR) and Unclassified Vessel(UN) using GMM-EM classifier with OLPP method from seven feature sets using ORCADES dataset.

Features		**OLPP**	
		C%	**UN%**
9	**GRB**	91.1	8.85
	GRH	91.7	11.34
	GRV	91.69	11.47
12	**GRBH**	90.68	10.35
15	**GRBHV**	91.13	11.5
	Mean (9-15 features)	**91.14**	**10.7**
	Standard Deviation (9-15 features)	**0.26**	**1.16**
32	**GRBHVL Contrast: GR**	90.1	9.2
37	**GRBHVLGy Contrast: GR**	90.21	10.85
42	**GRBHVL Profile: GR Contrast: GR**	90.4	8.98
47	**GRBHVLGy Profile:GR Contrast:GR**	90.56	8.85
	Mean (9-47 features)	**90.77**	**10.14**
	Standard Deviation (9-47 features)	**0.49**	**1.22**

Table 5 gives the performance measures, computed separately for arterioles and venules, of the system using OLPP with 12 feature set w.r.t. observer 1 on the ORCADES dataset. Table 6 presents the results obtained with OLPP using DRIVE dataset.

Figure 4 shows an example of an image with resulting vessel categorisation using OLPP with quadrant-pairwise GMM-EM classifier.

Table 5: Table showing the performance measures of OLPP with 12 feature set w.r.t. observer 1 on the ORCADES dataset.

Performance Measure	**OLPP**	
	arterioles	**venules**
Sensitivity	0.9618	0.8405
Specificity	0.8405	0.9618
Positive Predicted value	0.8791	0.9481
Negative Predicted value	0.9481	0.8791
Positive Likelihood Ratio	6.03	22.02
Negative Likelihood Ratio	0.045	0.1658

Table 6: Correct Classification Rate (CR) and Unclassified Vessel(UN) using GMM-EM classifier with OLPP on DRIVE dataset

Features		**OLPP**	
		C%	**UN%**
12	**GRBH Contrast: GR**	86.3	5.8
15	**GRBHV**	86.39	14
	Mean (12-15 features)	**86.34**	**9.9**
32	**GRBHVL Contrast: GR**	85.4	11.7
37	**GRBHVLGy Contrast: GR**	86.3	9.9
42	**GRBHVL Profile: GR Contrast: GR**	86.57	12.86
47	**GRBHVLGy Profile:GR Contrast:GR**	86.7	11.7
	Mean (12-47 features)	**86.3**	**10.9**

6 Discussion

As seen from Table 4, we observed that in smaller dimensions ($9-15$ features) as well as in larger dimension ($32-47$ features), the mean classification rates obtained are similar. We observed that the classification rate and hence the performance of the proposed method does not changes much irrespective of

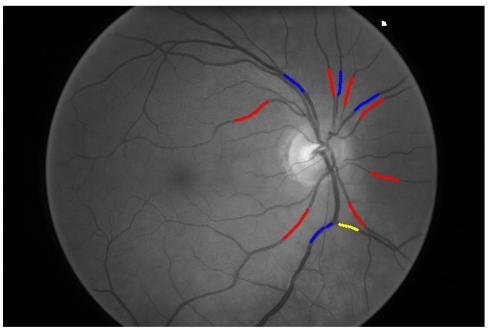


Fig. 4: ORCADES image showing the classification output using OLPP with quadrant-pairwise GMM-EM classifier in zone B. Vessels highlighted in red, blue and yellow have been classified as *arteriole*, *venule* and unclassified, respectively.

the feature dimension i.e. the proposed methods constantly perform well with all the feature subsets.

The performance measures using OLPP were computed with 12 feature set w.r.t. observer 1 on the ORCADES dataset (see Table 5). As seen from Table 5 the sensitivity for arterioles and venules, as obtained with OLPP is 0.9618 and 0.84505 respectively. That is to say, the probability of an incorrect classification is 3.8% and 15.9% for arterioles and venules respectively. The Positive predicted value (PPV) (precision) and Negative predicted value (NPV), for both arterioles and venules, with OLPP, are higher. Moreover, the high positive likelihood ratios (PLR) and low negative likelihood ratios (NLR) obtained with OLPP confirmed the high reliability of the system.

Additionally, the same observation, that OLPP consistently performed well, was observed when the analysis was performed on a DRIVE dataset (see Table 6).

Moreover, during the analysis, it was observed that correcting 7 channels to extract 47 features set was computationally expensive. In addition, further computational time is necessary for extraction of ROI based, profile-based and contrast features. Therefore, extracting large numbers of features from a retinal fundus image takes a significant amount of computational time. Hence, if the faster processing of large image datasets is desirable, OLPP is a promising method as the dimensionality of the feature set does not affect vessel classification accuracy. Moreover, this further gives more flexibility to the user to use any dimension of feature set to perform vessel classification.

The results showed that the proposed hybrid approach (OLPP+ GMM-EM) seems to be robust due the added advantage of GMM-EM and OLPP. With GMM-EM, the degree of uncertainty about the data point regarding its cluster assignment, make it possible to choose the cluster with highest

probability of assignment. Whereas, advantage of OLPP is that it has more locality preserving power.

Through analysis, we showed that the proposed methodology performed well in small as well as a large dimensional dataset. Thus, when compared with other conventional feature extraction method, such as PCA, the proposed method has an advantage that well-distributed classes in small datasets are formed thus giving the high-performance rate. Moreover, PCA takes into account the entire data for the principal components analysis without taking into consideration the fundamental class structure [42,43] whereas, as mentioned before, OLPP has a neighbourhood preserving property and can, therefore, better capture the intrinsic manifold structure to a greater extent.

7 Conclusions

We proposed a flexible approach for retinal vessel classification using OLPP, to automatically extract the most dominant and discriminating set for an image to then use with GMM-EM unsupervised classifier. Due to its ability to adapt to variations within and between images, the hybrid approach (OLPP + GMM-EM) gave high classification rate. Moreover, the proposed method found to be more consistent at different feature dimension. The proposed system has a broad application, for example, OLPP based feature extraction approach can apply to facial recognition and object identification solutions [44] and decision making to plan and organise service [45,46] during the Hajj and Umrah rituals. Moreover, the proposed method can be used to extract dominant features needed for Velocity-Based Modeling [47].

Additionally, in the field of securing data in Internet of Things (IoT) Health-Care Systems, the secure algorithm is required that will map best features of lightweight symmetric and asymmetric algorithms. The lack of security features affects the patient security and privacy. The proposed OLPP method might contribute in extracting the best features for the encryption of the data [48–51]. Furthermore, the application of proposed method can also be applied to the field of steganography. The performance of steganalysis is greatly influenced by the selection of feature subspaces [52,53]. The classification accuracy reduces and computational complexity increases by using a large number of features for Steganalysis relative to the size of the training set [52,54–56]. Thus to select feature subspaces more effectively the proposed method can also be tested in the field of Steganography [54–56].

Although here we utilised colour features, in our future research work, we are working on to design a system which selects dominant features from even larger feature bank prior to classification in an unsupervised manner. Additional test with different features and classifier can be done as a future work to further evaluate the performance of OLPP on different datasets. In future, to improve the computational speed, multi-core techniques such as GPU computing [57] and optimisation computing proposed in [58] can be explored for analysing large-scale dataset for biomarker discovery studies. Further tests

with different datasets are needed featuring images with different resolutions and from different camera systems to declare its suitability to support a-v classification in biomarker research.

Acknowledgment

This work was supported by Leverhulme Trust grant RPG-419 "Discovery of retinal biomarkers for genetics with large cross-linked datasets". Support from NHS Lothian R&D and the Edinburgh Clinical Research Imaging Centre is gratefully acknowledged.

References

1. M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010.
2. M. L. Baker, P. J. Hand, J. J. Wang, and T. Y. Wong, "Retinal signs and stroke," *Stroke*, vol. 39, no. 4, pp. 1371–1379, 2008.
3. S. Frost, Y. Kanagasingam, H. Sohrabi, J. Vignarajan, P. Bourgeat, O. Salvado, V. Villemagne, C. Rowe, S. L. Macaulay, C. Szoeke, *et al.*, "Retinal vascular biomarkers for early detection and monitoring of alzheimer's disease," *Translational psychiatry*, vol. 3, no. 2, p. e233, 2013.
4. M. Niemeijer, B. van Ginneken, and M. D. Abràmoff, "Automatic classification of retinal vessels into arteries and veins," in *SPIE medical imaging*, pp. 72601F–72601F, International Society for Optics and Photonics, 2009.
5. M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. van Ginneken, J. C. Folk, and M. D. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs.," *IEEE transactions on medical imaging*, vol. 30, pp. 1941–50, Nov. 2011.
6. A. Zamperini, A. Giachetti, E. Trucco, and K. S. Chin, "Effective features for artery-vein classification in digital fundus images," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, pp. 1–6, IEEE, 2012.
7. C. Muramatsu, Y. Hatanaka, T. Iwase, T. Hara, and H. Fujita, "Automated selection of major arteries and veins for measurement of arteriolar-to-venular diameter ratio on retinal fundus images.," *Computerized medical imaging and graphics*, vol. 35, pp. 472–80, Sept. 2011.
8. C. Kondermann, D. Kondermann, and M. Yan, "Blood vessel classification into arteries and veins in retinal images," in *Medical Imaging*, pp. 651247–651247, International Society for Optics and Photonics, 2007.
9. Q. Mirsharif, F. Tajeripour, and H. Pourreza, "Automated characterization of blood vessels as arteries and veins in retinal images," *Computerized Medical Imaging and Graphics*, vol. 37, no. 7, pp. 607–617, 2013.
10. B. Dashtbozorg, A. M. Mendonça, and A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1073–1083, 2014.
11. X. Xu, W. Ding, M. D. Abràmoff, and R. Cao, "An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image," *Computer Methods and Programs in Biomedicine*, vol. 141, pp. 3–9, 2017.
12. V. Vijayakumar, D. D. Koozekanani, R. White, J. Kohler, S. Roychowdhury, and K. K. Parhi, "Artery/vein classification of retinal blood vessels using feature selection," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pp. 1320–1323, IEEE, 2016.
13. N. Hatami and M. Goldbaum, "Automatic identification of retinal arteries and veins in fundus images using local binary patterns," *arXiv preprint arXiv:1605.00763*, 2016.

14. E. Grisan and A. Ruggeri, "A divide et impera strategy for automatic classification of retinal vessels into arteries and veins," in *Engineering in medicine and biology society, 2003. Proceedings of the 25th annual international conference of the IEEE*, vol. 1, pp. 890–893, IEEE, 2003.
15. M. Saez, S. González-Vázquez, M. González-Penedo, M. A. Barceló, M. Pena-Seijo, G. C. de Tuero, and A. Pose-Reino, "Development of an automated system to classify retinal vessels into arteries and veins," *Computer methods and programs in biomedicine*, vol. 108, no. 1, pp. 367–376, 2012.
16. V. S. Joshi, M. K. Garvin, J. M. Reinhardt, and M. D. Abramoff, "Automated artery-venous classification of retinal blood vessels based on structural mapping method," *Proc. SPIE Medical Imaging, Computer-Aided Diagnosis*, vol. 8315, p. 83150I, 2012.
17. G. Hamednejad and H. Pourghassem, "Retinal blood vessel classification based on color and directional features in fundus images," in *Biomedical Engineering (ICBME), 2015 22nd Iranian Conference on*, pp. 257–262, IEEE, 2015.
18. P. Pudil, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, pp. 1119–1125, 1994.
19. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
20. Y. Chen, F. He, Y. Wu, and N. Hou, "A local start search algorithm to compute exact hausdorff distance for arbitrary point sets," *Pattern Recognition*, vol. 67, pp. 139–148, 2017.
21. S. Piramuthu, "The hausdorff distance measure for feature selection in learning applications," in *Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on*, pp. 6–pp, IEEE, 1999.
22. E. P. Ong, J. A. Lee, G. Xu, B. H. Lee, and D. W. Wong, "An automatic quantitative measurement method for performance assessment of retina image registration algorithms," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pp. 3252–3255, IEEE, 2016.
23. E. P. Ong, Y. Xu, D. W. K. Wong, and J. Liu, "Retina verification using a combined points and edges approach," in *Image Processing (ICIP), 2015 IEEE International Conference on*, pp. 2720–2724, IEEE, 2015.
24. K. Li, F.-z. He, H.-p. Yu, and X. Chen, "A correlative classifiers approach based on particle filter and sample set for tracking occluded target," *Applied Mathematics-A Journal of Chinese Universities*, vol. 32, no. 3, pp. 294–312, 2017.
25. M. Miri, Z. Amini, H. Rabbani, and R. Kafieh, "A comprehensive study of retinal vessel classification methods in fundus images," *Journal of medical signals and sensors*, vol. 7, no. 2, p. 59, 2017.
26. K. C. Jordan, M. Menolotto, N. M. Bolster, I. A. Livingstone, and M. E. Giardini, "A review of feature-based retinal image analysis," *Expert Review of Ophthalmology*, vol. 12, no. 3, pp. 207–220, 2017.
27. D. Cai and X. He, "Orthogonal locality preserving indexing," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 3–10, ACM, 2005.
28. R. McQuillan, A.-L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, *et al.*, "Runs of homozygosity in european populations," *The American Journal of Human Genetics*, vol. 83, no. 3, pp. 359–372, 2008.
29. J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE transactions on medical imaging*, vol. 23, no. 4, pp. 501–509, 2004.
30. D. Relan, T. MacGillivray, L. Ballerini, and E. Trucco, "Retinal vessel classification: sorting arteries and veins," *35th Annual International Conference of the IEEE EMBS Engineering in Medicine and Biology Society (EMBC), Osaka, Japan*, pp. 7396 –7399, July 2013.
31. R. Chrástek, M. Wolf, K. Donath, H. Niemann, D. Paulus, T. Hothorn, B. Lausen, R. Lämmer, C. Y. Mardin, and G. Michelson, "Automated segmentation of the optic nerve head for diagnosis of glaucoma.," *Medical image analysis*, vol. 9, pp. 297–314, Aug. 2005.

32. A. Giachetti, K. S. Chin, E. Trucco, C. Cobb, and P. J. Wilson, "Multiresolution localization and segmentation of the optical disc in fundus images using inpainted background and vessel information," in *2011 18th IEEE International Conference on Image Processing*, pp. 2145–2148, Sept 2011.
33. J. Canny, "A computational approach to edge detection.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 8, pp. 679–98, July 1986.
34. S. Vazquez, B. Cancela, N. Barreira, M. G. Penedo, and M. Saez, "On the automatic computation of the arterio-venous ratio in retinal images: Using minimal paths for the artery/vein classification," in *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pp. 599–604, IEEE, 2010.
35. H. Jelinek, C. Depardieu, C. Lucas, D. Cornforth, W. Huang, and M. Cree, "Towards vessel characterisation in the vicinity of the optic disc in digital retinal images," in *Image Vis Comput Conf*, pp. 2–7, Citeseer, 2005.
36. X. He, D. Cai, H. Liu, and W.-Y. Ma, "Locality preserving indexing for document representation," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 96–103, ACM, 2004.
37. B. Tang, F. Li, and Y. Qin, "Fault diagnosis model based on feature compression with orthogonal locality preserving projection," *Chin J Mech Eng*, vol. 24, no. 5, pp. 891–898, 2011.
38. J. W. Yoon, "An efficient model selection for gaussian mixture model in a bayesian framework," *arXiv preprint arXiv:1307.0995*, 2013.
39. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
40. "Gaussian mixture model." `http://home.deib.polimi.it/matteucc/Clustering/tutorial-html/mixture.html`.
41. J. Leskovec, A. Rajaraman, and J. D. Ullman, *Dimensionality Reduction*. Cambridge University Press, 2014.
42. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, vol. 1, pp. 947–954, IEEE, 2005.
43. C. Li, Y. Diao, H. Ma, and Y. Li, "A statistical pca method for face recognition," in *Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium on*, vol. 3, pp. 376–380, IEEE, 2008.
44. S. Aly and A. Gutub, "Intelligent recognition system for identifying items and pilgrims," *NED University Journal of Research*, vol. Thematic Issue on Advances in Image and Video Processing, pp. 17–23, 2018.
45. N. Alharthi and A. Gutub, "Data visualization to explore improving decision-making within hajj services," *Sci Modell Res*, vol. 2, no. 1, pp. 9–18, 2017.
46. A. Gutub and N. Alharthi, "Improving hajj and umrah services utilizing exploratory data visualization techniques," in *16th scientific Hajj research Forum, Organized by the Custodian of the Two Holy Mosques Institute for Hajj Research, Umm Al-Qura University - King Abdulaziz Historical Hall, Makkah, Saudi Arabia*, May 2016.
47. S. Kim, S. J. Guy, K. Hillesland, B. Zafar, A. A.-A. Gutub, and D. Manocha, "Velocity-based modeling of physical interactions in dense crowds," *The Visual Computer*, vol. 31, no. 5, pp. 541–555, 2015.
48. S. Sangariand and L. Manickam, "A light-weight cryptography analysis for wireless based healthcare applications," *Journal of Computer Science*, vol. 10, no. 10, pp. 2088–2094, 2014.
49. A. K. Luhach *et al.*, "Analysis of lightweight cryptographic solutions for internet of things," *Indian Journal of Science and Technology*, vol. 9, no. 28, 2016.
50. N. Alassaf, B. Alkazemi, and A. Gutub, "Applicable light-weight cryptography to secure medical data in iot systems," *Journal of Research in Engineering and Applied Sciences (JREAS)*, vol. 2, pp. 50–58, 04 2017.
51. A. Gutub, N. Al-Juaid, and E. Khan, "Counting-based secret sharing technique for multimedia applications," *Multimedia Tools and Applications*, pp. 1–29, 2017.
52. C. Yang, Y. Zhang, P. Wang, X. Luo, F. Liu, and J. Lu, "Steganalysis feature subspace selection based on fisher criterion," pp. 514–521, 10 2017.

53. A. Adlakha and R. R. Chhikara, "Comparative analysis of filter feature selection techniques with different classifiers for image steganalysis," in *Computing, Communication and Automation (ICCCA), 2016 International Conference on*, pp. 1122–1127, IEEE, 2016.
54. A. Gutub and N. Aljuaid, "Multi-bits stego-system for hiding text in multimedia images based on user security priority," *Journal of Computer Hardware Engineering*, 04 2018.
55. A. A. G. Nouf A. Al-Juaid and E. A. Khan, "Enhancing pc data security via combining rsa cryptography and video based steganography," *Journal of Information Security and Cybercrimes Research (JISCR)*, vol. 1, no. 1, pp. 8–18, 2018.
56. N. A. Al-Otaibi and A. A. Gutub, "2-leyer security system for hiding sensitive text data on personal computers," *Lecture Notes on Information Theory*, vol. 2, no. 2, pp. 151–157, 2014.
57. Y. Zhou, F. He, N. Hou, and Y. Qiu, "Parallel ant colony optimization on multi-core simd cpus," *Future Generation Computer Systems*, vol. 79, pp. 473–487, 2018.
58. X. Yan, F. He, N. Hou, and H. Ai, "An efficient particle swarm optimization for large-scale hardware/software co-design system," *International Journal of Cooperative Information Systems*, vol. 27, no. 01, p. 1741001, 2018.